

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: August 17, 2005, 22:34:15 ; Search time 7192 Seconds  
(without alignments)  
7351.400 Million cell updates/sec

Title: US-10-695-243-3

Sequence: 1 atgcgaataatcgaagaag.....aacagcagcgtctcataa 1389

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database: EST.\*  
Listing first 45 summaries

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hc1.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gse1.\*  
9: gb\_gse2.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1389	100.0	3404 3	AK030141 Mus muscu
2	778.4	56.0	783 7	CK638784 UI-M-HO-
3	753.4	54.2	757 6	CB722971 UI-M-GH-
4	713.4	51.4	726 6	CB722993 UI-M-GH-
5	711.6	51.2	719 7	CO431563 UI-M-HX-
6	706.4	50.9	717 7	CK639206 UI-M-HO-
7	704.4	50.7	737 7	CN536849 UI-M-HO-
8	669.8	48.2	818 7	CN537400 UI-M-HO-
9	666	47.9	681 7	CN533999 UI-M-HO-
10	621.8	44.8	786 7	CN525978 UI-M-HO-
11	596	42.9	702 7	CF728873 UI-M-HO-
12	587.4	42.3	589 6	CA878786 B616889
13	557.6	40.1	668 2	BB616889 B616889
14	520.4	37.5	633 9	CG589448 OST241683
15	512	36.9	863 1	AJ813286 AJ813286
16	502.4	36.2	504 6	CA878760 K0969C10-
17	495.2	35.7	778 4	BG194979 RST14037
18	494.4	35.6	778 4	CA455409 UI-M-HO-
19	480	34.6	480 9	CF907159 A0507A10-
20	449	32.3	739 9	AG437766 Mus muscu
21	425	30.6	840 4	BG286536 602382535
22	424.6	30.6	487 6	CB726469 MAGNNTC:C
23	420.6	30.3	765 7	CK774533 963302 MA
24	414.2	29.8	591 4	BI848382 470812 MA

25	394	28.4	634	4	BM729307	UI-E-E01-
26	391	28.1	460	7	CK344479	K0955A10-
27	372.4	26.8	632	4	BM781108	MLN1_4_E0
28	369.4	26.6	651	5	BK471708	DKF2P686B
29	365.4	26.3	678	1	AL710783	DKF2P686E
30	361	26.0	614	5	BH218698	BU218698 603756354
31	357.6	25.7	799	5	BU365777	603784922
32	351.2	25.3	824	6	AY412826	Mus muscu
33	350	25.2	576	6	CD683768	EST288 hu
34	343.4	24.7	583	5	BK472485	DKF2P686J
35	341.6	24.6	824	9	AY412824	Homo sapi
36	340.2	24.5	752	7	CK596565	AGENCOURT
37	336.8	24.2	811	9	AY412825	pan trogl
38	332.2	23.7	609	7	CR753261	DKF2P463K
39	318.2	22.9	637	5	BU346013	604168885
40	305	22.0	610	5	BK481372	DKF2P686M
41	280	20.2	280	2	BF151483	u21510.Y
42	276.4	19.9	586	5	BP326282	BP326282
43	275	19.8	790	4	BG386631	602454883
44	274.8	19.8	827	5	BU386802	603581728
45	273.4	19.7	521	7	CN277241	170006001

## ALIGNMENTS

RESULT 1  
AK030141  
LOCUS  
DEFINITION  
Mus musculus adult testis cDNA, RIKEN full-length enriched library, clone:4932703L02 product:ENNO-ALPHA-D-MANNOSIDASE homolog (Rattus norvegicus), full insert sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AK030141.1 GI:26326122  
HTC; CAP trapper.  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED

1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636

2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159

3  
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishitani, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Karai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

BM729307 UI-E-E01-  
CK344479 K0955A10-  
BM781108 MLN1\_4\_E0  
BK471708 DKF2P686B  
AL710783 DKF2P686E  
BU218698 603756354  
BU365777 603784922  
AY412826 Mus muscu  
CD683768 EST288 hu  
BK472485 DKF2P686J  
AY412824 Homo sapi  
CK596565 AGENCOURT  
AY412825 pan trogl  
CR753261 DKF2P463K  
BU346013 604168885  
BK481372 DKF2P686M  
BF151483 u21510.Y  
BP326282 BP326282  
BG386631 602454883  
BU386802 603581728  
CN277241 170006001

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 18:09:50 ; Search time 10073 Seconds  
(without alignment)  
6681.653 Million cell updates/sec

Title: US-10-695-243-1

Perfect score: 1389  
Sequence: 1 atggcaagattcgagagaag.....gccagctgcctcttctta 1389

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl1:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1389	100.0	1389	6	CQ868818 Sequence
2	1387.4	99.9	4386	6	CQ491652 Sequence
3	1387.4	99.9	4386	6	CQ497537 Sequence
4	1387.4	99.9	4596	6	HSMB06986 Sequence
5	1385.8	99.8	1474	9	HSAS77574 Homo sapi
6	1385.8	99.8	4558	9	HSMB05695 Sequence
7	1075	77.4	2388	6	BD159701 Homo sapi
8	1075	77.4	2388	6	AX882397 Sequence
9	1075	77.4	2388	6	AK022900 Homo sapi
10	1065.8	76.7	4645	10	BC067076 Mus muscu
11	1064.2	76.6	1389	6	CQ868820 Sequence
12	929.4	50.9	2552	10	AF023657 Rattus no
13	706.8	47.3	105273	5	CR354351 Gallus ga
14	656.8	46.1	2779	5	BC077304 Xenopus la
15	640.8	43.1	782	9	AY048774 Homo sapi
16	599	39.5	1096	9	AY048775 Homo sapi
17	549.2	37.0	193704	10	AL805949 Mouse DNA
18	514.2	36.9	713	6	CQ482870 Sequence
19	513.2				

20	509.4	36.7	256771	2	AC115326	AC115326 Rattus no
21	456.8	32.9	2770	10	BC065047	BC065047 Mus muscu
22	408.8	29.4	1083	6	CQ722353	CQ722353 Sequence
23	408.8	29.4	2310	6	AR541894	AR541894 Sequence
24	372	26.8	584	6	CQ503974	CQ503974 Sequence
25	372	26.8	584	6	CQ512775	CQ512775 Sequence
26	356	25.6	1859	9	BC063587	BC063587 Homo sapi
27	284	20.4	348	6	BD148022	BD148022 Homo sapi
28	284	20.4	348	6	AX887960	AX887960 Sequence
29	281.8	20.3	3286	5	BC081681	BC081681 Danio rer
30	277.6	20.0	197552	10	AL606933	AL606933 Mouse DNA
31	277	19.9	449	6	CQ473701	CQ473701 Sequence
32	270.4	19.5	797	11	BV174680	BV174680 sgmm73513
33	269.6	19.4	181070	2	AC142187	AC142187 Rattus no
34	264.6	19.0	123789	2	AC110674	AC110674 Tetradon
35	256	18.4	1246	5	BX930133	BX930133 Gallus ga
36	255.2	18.4	2341	6	AX714001	AX714001 Sequence
37	255.2	18.4	2341	9	AK055996	AK055996 Homo sapi
38	255.2	18.4	134161	9	AC104336	AC104336 Homo sapi
39	255.2	18.4	143060	9	AL928472	AL928472 Human DNA
40	255.2	18.4	170908	2	AC022557	AC022557 Homo sapi
41	255.2	18.4	190310	2	AC023225	AC023225 Homo sapi
42	255.2	18.4	255952	2	AL513473	AL513473 Homo sapi
43	209.8	15.1	1347	9	BC009952	BC009952 Homo sapi
44	193.2	13.9	301630	1	AE016942	AE016942 Bacteroid
45	178.6	12.9	67717	2	AC101393	AC101393 Mus muscu

#### ALIGNMENTS

RESULT 1  
CQ868818  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CQ868818  
Sequence 1 from Patent WO2004074497.  
CQ868818  
CQ868818.1 GI:51998752

Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1  
Hamilton, S.R.  
Endomannosidases in the modification of glycoproteins in eukaryotes  
Patent: WO 2004074497-A 1 02-SEP-2004;  
Hamilton, Stephen R. (US)

FEATURES  
source  
location/Qualifiers

1..1389

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

1..1389

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAH33902.1"

/db\_xref="GI:51998753"

/translation="MAKFRRTCTIILALFILFISLMGLKMLRPTATGAPFGDL  
LPELHRTYHLGKPNDRSDRINSETNKNKSVITTKPSASLNDLPLNNY  
LHVFPYSWYGNRQDPQKTYHNPVLEHNDPRAKYRPGGRNPPDIDSSPYELGS  
YSSRDPSTVETHRQKRSASIGTALSWTPDVNDNGEETDLVFTLDKAKYDK  
VYHILPYSNRDQNNYKAVKYLIDRYGNHAPVRYKRTGNALPMVFTYDSTYKPE  
KMANLITTSGRSIRNSPYDGLFIALIVEKHYDILQSGPDIYTPYFNGFTYSS  
HONWASIKLIDCKYNLIFIPVSGVGYDISIRPMWNTNRNRNGKYEYGLSALOT  
RPLSLITSPFENWHEGTQIEKAVPKRTSTVYLDVAPHKGLYELTRKMSKYSKER  
ATYALDRQLPVPS"

ORIGIN

Query Match 100.0%; Score 1389; DB 6; Length 1389;

Best local similarity 100.0%; Pred. No. 1.1e-298;

Matches 1389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGGCAAGTTTCGAGAGAGACTTGACATTCATTGGCACTTTTATTCATTATTATTTTC 60